Systems Engineering Roadmap for Dependable Autonomous Cyber-Physical Systems

Adler Rasmus
Program Manager Autonomous Systems
Fraunhofer IESE
Kaiserslautern, Germany
rasmus.adler@iese.fhg.de

***Abstract*—Autonomous cyber-physical systems have enormous potential to make our lives more sustainable, more comfortable, and more economical. Artificial Intelligence and connectivity enable autonomous behavior, but often stand in the way of market launch. Traditional engineering techniques are no longer sufficient to achieve the desired dependability; current legal and normative regulations are inappropriate or insufficient. This paper discusses these issues, proposes advanced systems engineering to overcome these issues, and provides a roadmap by structuring fields of action.**

***Keywords*—autonomous systems, systems engineering, dependability**

I. Introduction

A system can be described as autonomous if it is capable of independently achieving a predefined goal in accordance with the demands of the current situation without recourse to either human control or detailed programming [1]. Autonomous systems have enormous potential to contribute decisively to the solution of the current ecological, social, and economic challenges. The application fields are very broad and range from autonomous driving via autonomous manufacturing lines to medicine and nursing care. Autonomous systems make it possible to design road and rail transport in a more ecological, safer, and more efficient way. By using them, companies can make work and production processes more flexible and more efficient. In agriculture, they can lead to higher yields and, at the same time, reduced use of fertilizers, and can thus contribute to better alignment between ecology and economy. In the energy transition, energy management can be optimized with the help of the cognitive capabilities of autonomous systems. They can counteract the shortage of skilled workers in many areas such as construction, freight and passenger transport, medicine, and nursing care, despite the demographic change. Autonomous behavior is usually based on the four steps shown in Figure 1.

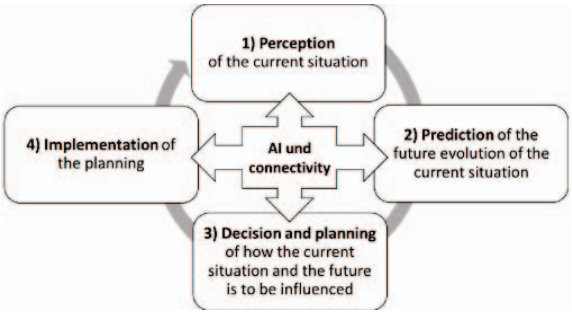

Figure 1 – Cognitive loop

This loop is also referred to as the cognitive loop, the "Sense Understand Decide Act (SUDA)" model [7], or the "Monitor Analyze Plan Execute – Knowledge (MAPE-K)" cycle [6].

If each of these steps in the loop is realized by detailed programming, we do not meet the above definition of autonomous behavior. If everything has been planned in advance, no matter how complicated it may be, and if we program these planned causal relationships into the system, we speak of an automated system. But if we do not really capture the causal relationships and if we use methods from the field of Artificial Intelligence (AI) to tell the system only indirectly how to behave in a particular situation, then we speak of an autonomous system. The most prominent example in the context of autonomous vehicles is object classification based on artificial neural networks. Object classification is an essential prerequisite for prediction because kinematic information about objects is typically not sufficient to predict possible accident scenarios or other scenarios accurately. For the prediction of trajectories themselves, AI methods like Bayesian networks can also be used in a meaningful way [4].

Connectivity also plays a crucial role in realizing the cognitive loop. For instance, it enables autonomous vehicles to share their knowledge about the current situation with other autonomous vehicles. Planned behavior can also be shared so that future scenarios can be anticipated better. Last but not least, behavior planning can transcend the level of a single vehicle and can be optimized by a network of vehicles with regard to the goals. Consequently, autonomous systems are often cyber-physical systems.

The application of AI and the cyber-aspect comes with many challenges for assuring dependability, and particularly for assuring safety. This paper shall contribute to a holistic engineering approach for dependable autonomous systems by structuring important fields of action and integrating some existing solutions.

The paper is structured as follows. Section II briefly discusses the current legal and normative situation with respect to safety assurance of autonomous machinery. One issue here is that the machinery directive and its related harmonized standard ISO 12100 focus on the engineering of the machinery and not the context in which the machinery is acting autonomously. In section III, we argue that this scope is too small for engineering dependable autonomous machinery because fundamental design decisions for avoiding and minimizing risks are hard to address. For instance, the complexity of the physical environment is typically reflected in the cognitive loop. This leads to a tradeoff between changing the environment to lower its complexity and engineering more intelligent autonomous machinery that can cope with a more complex environment. We propose an advanced systems engineering approach that reflects such fundamental design decisions in design-time models that are then transformed into AI-based runtime models for making the autonomous systems risk-aware and enable them to autonomously manage risks at runtime. Section IV summarizes and reflects the discussion on this holistic approach.

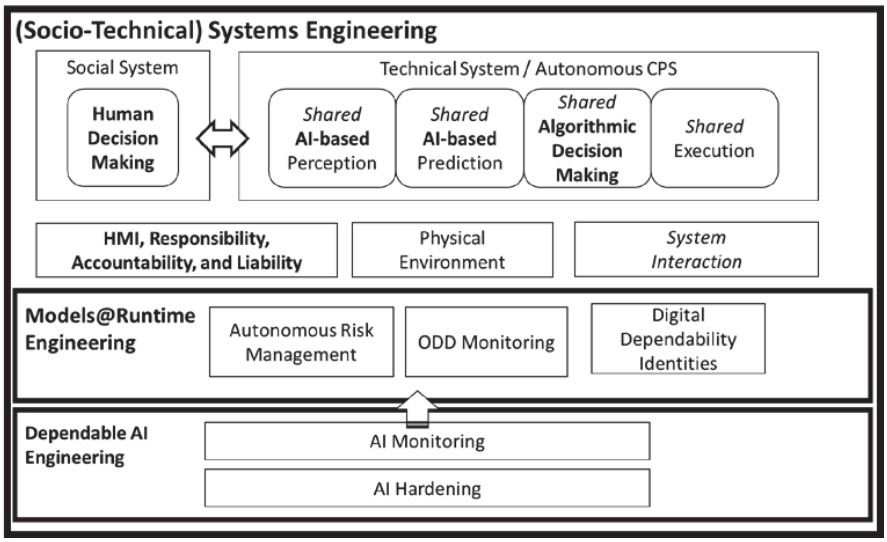

Figure 2 – Illustration of relevant engineering issues for dependable autonomous systems

II. Legal and Normative Aspects Regarding AI and Connectivity

The behavior of machinery is generally safety-critical and subject to legal and normative safety requirements. In terms of autonomous behavior based on dynamic connectivity and AI, this presents great challenges to manufacturers.

Current laws and standards are still very much oriented towards the physical limits of a machine and thus in conflict with connectivity. In Europe, machines must comply with the fundamental safety requirements of the EC Machinery Directive 2006/42/EG. The Machinery Directive defines a machine, among other things, as "an assembly of linked parts". DIN EN ISO 12100, which is harmonized with the Machinery Directive, similarly defines machinery as an "assembly of interconnected parts or sub-assemblies…". Laws and standards thus suggest implementing all four steps for autonomous behavior in the machinery because even if the control of the machinery is completely separated, the manufacturer of the machinery currently bears the main responsibility for its safe behavior. The use of AI for safety-critical control systems is also critical from a legal and normative perspective, as there is a lack of clear specifications. Neither the Machinery Directive nor the DIN EN ISO 12100 standard is dedicated to the development of software. For safety-critical software, IEC 61508 is primarily relevant. DIN EN ISO 12100 references this basic standard for functional safety. IEC 61508 is to be applied when an area of application is involved for which no specially derived standard exists. Derived standards include, in particular, ISO 13849, ISO 26262, and ISO 25119. Neither IEC 61508 nor any of its derived standards currently clarify under which circumstances which type of AI may be used for safety-critical control systems. This holds also for ISO/PAS 21448, which complements ISO 26262 with respect to "Safety Of the Intended Functionality (SOTIF)" including functional insufficiencies and reasonably foreseeable misuse.

Although many standards already explicitly address AI, they do not extend safety engineering. For example, in ISO/IEC TR 24028 "Overview of trustworthiness in artificial intelligence", the only statement about functional safety is that dedicated safety functions may be introduced, which should then be implemented in accordance with IEC 61508. The upcoming technical report ISO/IEC AWI TR 5469 "Artificial intelligence — Functional safety of AI-based systems" is being developed by safety and AI experts and aims to provide more guidance. It describes different classes of AI and usage contexts in order to determine for each combination whether traditional safety standards are sufficient, additional requirements apply, or no AI should be used.

In the context of the current revision of the Machinery Directive, adaptations regarding connectivity and AI are being discussed. On the normative level, there are also many efforts to adapt the safety requirements. The principal aim is to promote dependable innovations, but also to assure dependability. Dependability in this context encompasses especially safety and security. However, engineering must always consider all quality attributes and find an acceptable compromise. Standards, in particular, serve to define what is acceptable. In this respect, it must be considered what is currently possible from an engineering perspective, but also what is possible in the long term because laws, in particular, should be as stable as possible. The Machinery Directive was not significantly revised since 2006. If we aim at such a stable solution, then we should try to look ahead roughly until 2035.

In the project "AI Testing and Auditing" [2], Fraunhofer IESE is investigating together with legal experts whether or how the current regulations and standards should be revised. From a technical perspective, the basis for this is a commonly accepted dependability engineering roadmap and framework for autonomous cyber-physical machinery. To this end, the following section will discuss what such a roadmap and framework could look like.

III. Systems Engineering Roadmap

The safeTRANS Roadmap "Safety, Security, and Certifiability of Future Man-Machine Systems" describes how cyber-physical systems will evolve across domains until 2035 in terms of autonomy, intelligence, connectivity, and other aspects. The "man-machine" in the title already indicates that the scope of engineering should be the overall socio-technical system and not only the technical part. The engineering of socio-technical systems with respect to safety has a long history. System safety standards like MIL 882E and methods like STAMP/STPA [8] already provide some

guidance. We believe that this guidance needs to be enhanced with respect to the engineering of dependable autonomy of technical systems. Accordingly, we propose an advanced system engineering approach comprising additional engineering aspects, as depicted in Figure 2. Starting from the state of the practice, an engineering roadmap for each aspect or field of action is presented below.

A. Field of action: Systems Engineering

Systems engineering for machinery currently focuses on the engineering of individual machinery and the interaction of different engineering disciplines such as mechanical engineering, electrical engineering, and computer science.

With regard to dependable autonomous behavior, systems engineering must broaden its focus and consider the entire **socio-technical** system in which the autonomous behavior is embedded. This necessary extension is already reflected in the new cross-domain application rule VDE-AR-E 2842-61 "Development and trustworthiness of autonomous/cognitive systems". The Solution Level in Part 3 is dedicated to the embedding of an autonomous system into its socio-technical system. Regarding this embedding, the following three aspects are relevant:

1) HMI and Responsibility Ascription Engineering:

From a socio-technical perspective, the technical subsystem only performs a partial task and works together with humans to complete a large overall task. In this context, "performing" does not necessarily imply bearing responsibility. Humans can monitor systems or vice versa. When humans are to monitor autonomous behavior, it is often seen that they cannot live up to their responsibility. Accidents that occur when highway pilots are used are a good example of the fact that humans are often incapable of bearing the responsibility assigned to them. When distributing responsibility, systems engineering should not only consider that tolerance is generally lower for technical errors than for human drivers, but also that the overall accident risks should be minimized. The latter is reflected already in the second ethical rule for automated and connected vehicular traffic "…The licensing of automated systems is not justifiable unless it promises to produce at least a diminution in harm compared with human driving, in other words a positive balance of risks". However, in other application domains, the risk acceptance criteria of "positive risk balance" is not established.

We see a trend that the decision-making process in a socio-technical system moves from the social subsystem to the technical subsystem. This means that the algorithmic decision making (ADM) [12] increases. Further, interaction between the two subsystems becomes more complex so that is no longer possible to define a simple interface between these subsystems and engineer them separately [10]. The transition from traditionally programmed ADM to AI-based ADM may also cause responsibility and liability gaps [11]. Existing approaches like STAMP/STPA and their sociotechnical models of control (cf. page 82 in [8]) consider the role of regulatory frameworks but not the specific challenges that arise from AI and autonomous systems described in [11].

2) Physical Environment:

Major challenges in the implementation of autonomous behavior are the perception of the current situation and the prediction of possible future scenarios. A sub-task of systems engineering is to simplify perception and prediction appropriately by limiting the situation space. By reducing the complexity of the physical environment, the autonomous behavior becomes inherently more dependable – but only if the simplified environmental conditions are ensured permanently. An extreme example with respect to autonomous trucks would be to build completely straight motorways without any curves in order to reduce the complexity of lateral control. Depending on the application domain and the organizations involved in the development and operation of autonomous systems, it is challenging to find an organization that can make such fundamental design decisions. For instance, the manufacturer of a truck is traditionally not involved in motorway construction. Designing the physical environment requires finding suitable business and operator models. Operator models such as those used in rail transport, where the costs for the infrastructure can be passed on to the train ticket, are not established yet in many sectors, including agriculture.

3) System Interaction:

As shown in Figure 2 by "Shared…", connectivity and interaction play a crucial role for autonomous behavior. First of all, it must be ensured that only reliable groups of machinery are formed. In addition, changing environmental conditions must not impair the dependability once a group has been formed. Conceptual solutions such as Digital Dependability Identities (DDI) [3] already exist to address this issue. DDIs have been tested in applications such as truck platooning [5], where trucks driving closely behind each other on the motorway exploit slipstream effects to save fuel. General challenges in this context are the standardization of the interaction protocols and the lack of added value if the number of machines capable of interacting is too small.

These three aspects are of crucial importance for engineering an inherently dependable solution. Furthermore, the engineering itself is not trivial because there are strong dependencies between these aspects, and the initial approach for each aspect might turn out to be infeasible, impractical, or too expensive. We therefore argue that established system engineering approaches should be enhanced with respect to holistic treatment of these aspects. Moreover, the models used for describing these three aspects are not only the basis for important decisions in systems engineering, but also for the second field of action, namely models@runtime.

B. Field of action: Models@Runtime

In practice, model-driven development still focuses on the use of models for design, implementation, or verification.

To make autonomous machinery dependable, runtime models must be increasingly used to monitor and verify runtime behavior or to implement "runtime certification". The following examples of runtime models result from the three aspects of systems engineering stated above.

1) ODD Monitoring

The first runtime model refers to the monitoring of the boundaries within which autonomous behavior is dependable. These boundaries primarily refer to the physical environment, but also to the other two aspects. In the automotive sector, the "Operational Design Domain (ODD)" is already a fixed term for the boundaries of dependable behavior. The ODD is primarily a design-time model that various stakeholders such as engineers, operators, and lawyers need for various

purposes. Accordingly, it is reasonable to present it in various ways that meet the requirements of the stakeholders. However, assuring that the behavior actually stays within the defined limits, it is generally necessary to synthesize a related runtime model for monitoring the ODD boundaries.

2) Autonomous Risk Management

A related runtime model refers to the autonomous management of the risks for which the system is responsible within the defined boundaries. This concerns, for instance, risks due to collisions. According to the concept of separation of concerns, an extra cognitive loop could achieve situational risk awareness and overwrite the cognitive loop for the nominal behavior. In the automotive sector, important work in this direction is provided, for instance, by the SINADRA approach [4].

3) Digital Dependability Identities (DDIs)

Finally, runtime models are used in the design of dynamic systems of systems. Digital twins are a prime example of runtime models. A special type of digital twins are DDIs [3]. They can be used to check at runtime whether systems that come together at runtime can cooperate dependably. If this is the case, they can be used to monitor the runtime cooperation and adapt it such that it will remain dependable. One application example for DDIs is truck platooning. DDIs make it possible to check first which vehicles are permitted at all to form a platoon and what distance they must keep due to static conditions such as the design of the braking system. In addition, they make it possible to minimize the distance depending on the situation. Instead of always assuming the worst case that applies to all situations, different cases are considered and corresponding solution variants are implemented. The assumptions under which a variant is safe are recorded in the DDIs. At runtime, these assumptions are continually reviewed and the appropriate variant is selected. As runtime models generally involve the usage of AI methods, and as AI is one of the core challenges for assuring the dependability of an autonomous system, dependable AI is another important field of action.

C. Field of action: Artificial Intelligence

Considering autonomous machinery, AI is currently primarily concerned with supervised learning and data-driven models. The related research field is very extensive. Regarding the models@runtime discussed above, it can be divided into two areas:

1) AI Hardening

On the one hand, the aim is to avoid defects in the model at development time, or to find and eliminate them. Fault avoidance and fault removal are two well-known means for dependability described in [13]. The basic concepts and the taxonomy in [13] are transferrable to AI but they have to be implemented in a different way. Defects can be avoided, for example, by providing suitable data for learning. Research areas such as Explainable AI provide methods for finding defects and thus subsequently improving the models. Despite intensive research, however, no approach has been found yet to make learned models as dependable as traditional software models. The term “defects” is used here because errors are often understood as a deviation from an explicit specification and because such a specification is replaced by data and a learning approach.

2) AI Monitoring

Complementary approaches to fault avoidance and removal are fault tolerance and forecasting [13]. Traditional approaches are, however, often not applicable as it is not only hard to specify what is intended but also what is critical. Novel approaches thus try to estimate the “uncertainty” of the output of a data-driven model. As depicted in Figure 2, this uncertainty estimation can be performed at runtime and considered in the aforementioned runtime models. For instance, autonomous risk management can increase the estimated risk if the uncertainty of the variables in the risk calculation formula increases.

IV. Summary and Reflection

In order to fully exploit the potential of autonomous commercial vehicles and benefit from autonomous behavior even in critical applications, engineering must change. A key success factor is to expand the scope of systems engineering. Depending on the application, the physical environment and the interaction with humans as well as with other systems must be considered to a much greater extent. There are already many technological solution components such as DDIs to enable cross-manufacturer innovations, but systems engineering remains the key to minimizing the risks and balancing them against the opportunities.

V. References

[1] Fachforum Autonome Systeme im Hightech-Forum: Autonome Systeme – Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft. Langversion, Abschlussbericht, Berlin, April 2017.

[2] EXAMAI, https://testing-ai.gi.de/, accessed on 09-10-2020

[3] DEIS project, https://www.deis-project.eu/, accessed on 09-10-2020

[4] Reich, Jan & Trapp, Mario. (2020). SINADRA: Towards a Framework for Assurable Situation-Aware Dynamic Risk Assessment of Autonomous Vehicles. 10.13140/RG.2.2.26063.71849.

[5] Reich J. et al. (2020) Engineering of Runtime Safety Monitors for Cyber-Physical Systems with Digital Dependability Identities. In: Computer Safety, Reliability, and Security. SAFECOMP 2020. Lecture Notes in Computer Science, vol 12234. Springer, Cham. https://doi.org/10.1007/978-3-030-54549-9_1

[6] J. O. Kephart and D. M. Chess. The vision of autonomic computing. IEEE Computer, 36(1):41–50, 2003

[7] Robust and resilient. Designing safe automated driving systems; https://www.york.ac.uk/assuring-autonomy/news/blog/safety-highly-automated-driving-robust-resilient/, accessed on 09-10-2020

[8] Leveson, Nancy G.. “Engineering a Safer World: Systems Thinking Applied to Safety.” (2012).

[9] BMVI 2017, Ethics Commission https://www.bmvi.de/SharedDocs/EN/Documents/G/ethic-commission-report.pdf?__blob=publicationFile, accessed on 09-10-2020

[10] Kyle J. Behymer, John M. Flach, From Autonomous Systems to Sociotechnical Systems: Designing Effective Collaborations, She Ji: The Journal of Design, Economics, and Innovation, Volume 2, Issue 2,2016, Pages 105-114

[11] Burton, Simon, Habli, Ibrahim , Lawton, Tom et al. (3 more authors) (2019) Mind the Gaps: Assuring the Safety of Autonomous Systems from an Engineering, Ethical, and Legal Perspective. Artificial Intelligence. 103201. ISSN 0004-3702

[12] Understanding algorithmic decision-making: Opportunities and challenges,https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf , accessed on 09-10-2020

[13] A. Avizienis, J. -. Laprie, B. Randell and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," in *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 1, pp. 11-33, Jan.-March 2004, doi: 10.1109/TDSC.2004